

CrossFlow: Interpolando Dados Pluviométricos com Apoio de Validação Cruzada em Workflows Científicos

**Ulisses Roque Tomaz¹, Ednaldo O. Santos³, Gustavo B. Lyra³,
Sergio Manuel Serra da Cruz^{1,2}**

¹ Programa de Pós-Graduação em Matemática Computacional - PPGMMC/UFRRJ

² Programa de Educação Tutorial - PET-SI/UFRRJ

³ Departamento Ciências Ambientais / Instituto de Florestas - UFRRJ

ulisses.rtomaz@ufrrj.br, serra@pet-si.ufrrj.br

Resumo. *Os estudos de eventos atmosféricos extremos são importantes para a sociedade em geral. A chuva, tendo a altura pluviométrica como principal variável, insere-se neste contexto. Os dados pluviométricos brutos muitas vezes apresentam-se como longas séries que contém erros e falhas. Essas condições representam um desafio para a análise de padrões e previsão de eventos. Esse trabalho apresenta uma abordagem inédita baseada em workflows científicos que conjugam quatro métodos de interpolação e de cruzamento de dados para o preenchimento de falhas nas séries históricas. Nossos experimentos utilizaram dezenas de estações no estado do Rio de Janeiro em um período de 75 anos e produziram análises e dados curados de alta qualidade e livres de falhas.*

Abstract. *The study of extreme events and atmospheric patterns is important to society. The rain, having its height as main studied variable, belongs to this context. Rainfall data are delivered as long historical series not errors-free. This work presents a novel approach based on scientific workflows, cross validation and the usage of four data interpolation methods for the gap filling in the historical series. Our experiments used dozens of stations in the state of Rio de Janeiro over a period of 75 years producing high quality cured data free of gaps.*

1. Introdução

A compreensão das mudanças climáticas é um dos principais temas de pesquisas da atualidade. Os modelos climáticos se tornaram uma das principais ferramentas para avaliar essas mudanças. No entanto, obter e processar grandes volumes de dados meteorológicos de qualidade e livres de falhas representa um grande desafio científico e tecnológico. Estudos de padrões e probabilidades de ocorrência de eventos meteorológicos extremos (inundações e secas) sempre foram importantes para a humanidade. Porém, em decorrência de mudanças climáticas ocorridas nas últimas décadas, a frequência desses eventos tem aumentado (MARENGO, 2009). Neste cenário, a chuva, dentre os principais fenômenos meteorológicos, é de maior interesse em estudos climáticos, pois constitui uma das etapas do ciclo hidrológico que ao longo dos anos sofre alterações.

Estudos nas áreas Ambientais, Meteorologia e Hidrologia requerem dados heterogêneos e livres de falhas (dados ausentes e/ou errôneos). Essas características, por si só, incrementam a complexidade da análise dos dados que, demandam grande esforço computacional para produzir resultados confiáveis e em tempo hábil, que permitem que os

cientistas compreendam os fenômenos ou mesmo que a sociedade se previna ou mitigue as consequências de eventos extremos. Dessa forma, coletar, integrar e analisar dados meteorológicos e hidrológicos curados são tarefas que requerem uso intensivo de diversas técnicas computacionais.

Os meteorologistas e hidrólogos trabalham em pequenos grupos e aplicam modelos complexos que utilizam numerosas fontes de dados de natureza heterogênea e distribuída (PLALE, 2011). Em geral, as etapas de trabalho executadas podem ser resumidas na Figura 1: i) coleta de dados; ii) limpeza de dados brutos; iii) assimilação de dados e execução de modelos e iv) análise de dados usando pacotes estatísticos ou sistemas de informação geográfica (ArcGIS, GRASS entre outros).

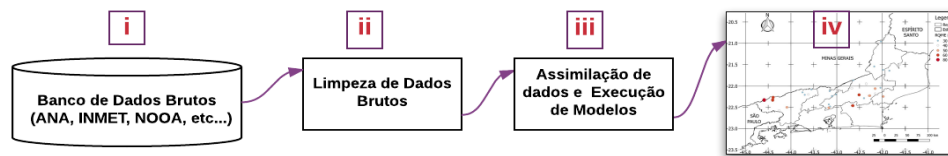


Figura 1 – Representação conceitual das etapas dos estudos em Meteorologia.

Segundo Plale (2011) um dos paradigmas computacionais mais apropriados para este tipo de estudos são os *workflows* científicos. Eles são capazes de modelar experimentos em larga escala, que encadeiam atividades e podem auxiliar tanto na rastreabilidade dos dados e processos quanto na reprodutibilidade dos experimentos e, assim, asseguram a transparência da metodologia adotada pelas equipes de pesquisas.

Portanto, o objetivo deste trabalho é apresentar uma abordagem sistemática e automática de coleta de dados brutos, preenchimento de falhas de dados pluviométricos mensais baseado em múltiplos métodos de interpolação apoiados por *workflows* científicos. A abordagem proposta, diferentemente das tradicionais que utilizam apenas um método de interpolação de dados, aplica concomitantemente quatro métodos de interpolação já consagrados na área da Meteorologia. Além disso, utiliza validação cruzada como critério de avaliação e seleção do melhor método de interpolação para estimar a altura pluviométrica adequada de dezenas de estações.

Esse artigo está organizado, além da Introdução, da seguinte forma: Seção 2 apresenta o referencial teórico; Seção 3 contém a abordagem denominada *CrossFlow*; Seção 4 apresenta as avaliações experimentais; Seção 5 discute os trabalhos relacionados. A Seção 6 apresenta as considerações finais destacando as contribuições e possíveis trabalhos futuros.

2. Referencial Teórico

O controle de qualidade de dados de precipitação não é uma tarefa trivial, no entanto, é fundamental para o sucesso de diversos tipos de estudos ambientais. A análise de consistência dos dados tem como objetivo garantir a qualidade dos dados pluviométricos por meio da identificação, correção de erros e preenchimento automático das falhas das séries de dados (SILVA *et al.*, 2014). Esses dados são essenciais para avaliar a altura pluviométrica, que representa a lâmina de água (em mm), que se formaria sobre o solo como resultado da chuva, caso a superfície fosse impermeável e não ocorresse escoamento ou evaporação.

A distribuição espaço-temporal das chuvas requer a análise de longas séries de dados e de conjuntos de estações pluviométricas próximas para permitir o preenchimento de lacunas nos registros ou substituição de dados observados e considerados errôneos. Assim, devem-se

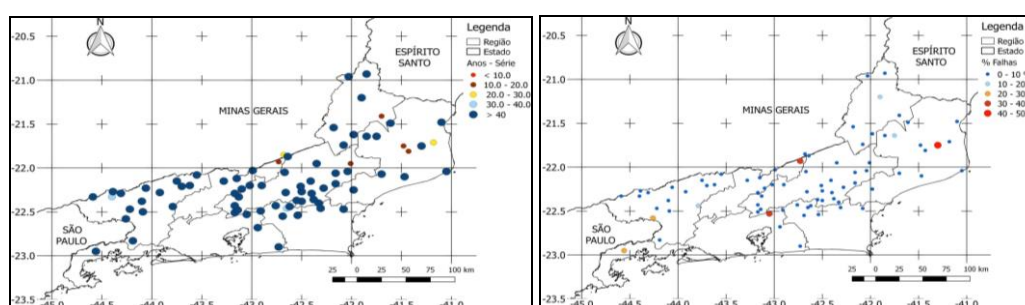
considerar estações situadas em uma mesma bacia hidrográfica ou em regiões vizinhas para a análise de registros pluviométricos (BERTONI & TUCCI, 2001). Adicionalmente, é importante que se obtenham da operadora da rede de estações informações sobre todas as irregularidades encontradas, a fim de melhorar a qualidade das séries de dados. Por fim, é desejável que o meteorologista conheça o regime climático, o sistema de circulação geral e processos geradores de chuvas, orografia, existência de microclimas e demais fatores que possam influenciar na ocorrência das chuvas na região em estudo (ANA, 2012).

Tradicionalmente, a análise de consistência possui duas fases distintas. Na *primeira fase* (análise preliminar) são verificadas as irregularidades na recepção dos dados pelos equipamentos de medição (avaliação dos dados diários e totais mensais, comparando-os com os das estações de apoio). Na *segunda fase* (análise quantitativa) são aplicados métodos de correção e homogeneização de dados pluviométricos.

2.1. Fase 1 - Análises Preliminares

As estações meteorológicas são analisadas, selecionadas e separadas em dois grupos: principais e secundárias. As principais são aquelas cujos registros em questão são obtidos e identificados a partir de características básicas, a saber: localização adequada na bacia para os propósitos do estudo; grandes séries temporais de observação (maior ou igual a 25 anos); poucos períodos de interrupção; presumivelmente operada adequadamente, com pouca ou nenhuma alteração na instalação (BERTONI & TUCCI, 2001; ANA, 2012).

As estações secundárias têm como finalidade verificar a existência de erros de observação ou transcrição nos dados das estações principais. Entretanto, em função dos resultados da análise de consistência, uma estação secundária pode ser levada à condição de principal pela qualidade dos seus registros e das suas condições naturais. Em geral, as estações secundárias devem ser selecionadas tendo como base os seguintes aspectos: maiores coeficientes de correlação (superior ou igual a 0,7); menores distâncias entre as estações (distâncias máximas de 200 km); períodos de dados coincidentes; e altitudes semelhantes entre as estações.



Figuras 2 - Distribuição espacial das 77 estações do estado do RJ considerando a quantidade de anos de observação das séries de dados (mapa à esquerda) e o percentual de falhas de preenchimento das estações (mapa à direita).

Neste trabalho, como resultado das análises preliminares, foram selecionadas 77 estações pluviométricas (principais e secundárias) da rede operada pela Agência Nacional de Águas (ANA) e pertencentes à Região Hidrográfica Atlântico Sudeste, que abrange parte das unidades hidrográficas Litorânea São Paulo/Rio de Janeiro, Paraíba do Sul e extremo sul da região Litorânea Rio de Janeiro/Espírito Santo, as coordenadas geográficas de latitude estão entre $20^{\circ} 93' S$ e $22^{\circ} 95' S$ e longitudes entre $41^{\circ} 85' W$ e $44^{\circ} 59' W$ (Figura 2).

2.2. Fase 2 - Análises Quantitativas

Dentre os métodos de interpolação para preenchimento de falhas de dados pluviométricos consagrados na literatura utilizaram-se neste trabalho: Regressão Linear (RL), Ponderação Regional (PR), Ponderação Regional com Regressão Linear (PRRL) e Inverso do Quadrado da Distância (IQD) (BERTONI & TUCCI, 2001; ANA, 2012). Além disso, neste trabalho usou-se como estratégia de validação cruzada (KOHAVI, 1995) o método *leave-one-out* para avaliar a capacidade de generalização da abordagem proposta. Neste método, utiliza-se uma única amostra para o subconjunto de teste, sendo as demais alocadas para o subconjunto de treinamento.

3. Abordagem Proposta

Nessa seção apresentamos a abordagem proposta, denominada *CrossFlow*. O *CrossFlow* é abordagem apoiada no paradigma dos *workflows* científicos, que podem ser acoplados a um SGWfC e dar suporte as análises preliminares e quantitativas (descritas na Seção 2), além de armazenar dados curados e de proveniência coletados pelo SGWfC. Uma representação abstrata da abordagem é apresentada na Figura 3. Em sua versão atual o *CrossFlow* foi codificado em Python para ser acoplado ao sistema *VisTrails* (CALLAHAN *et al.*, 2006).

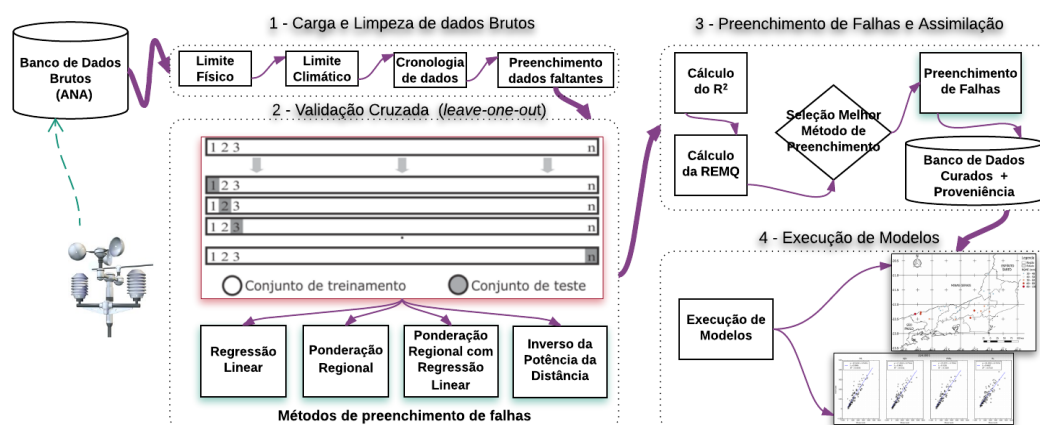


Figura 3 – Representação conceitual da abordagem **CrossFlow**.

O *CrossFlow* é composto por dois tipos de módulos principais: os que processam dados pluviométricos e os que fazem a conexão com as bases de dados. Os módulos de conexão são responsáveis pelo registro dos dados brutos e de proveniência retrospectiva no banco de dados. Nesta abordagem, toda a base de dados se encontra representada em um único banco de dados relacional no MySQL, pois é necessário manter a compatibilidade com os sistemas previamente desenvolvidos, *e.g.* sistemas *Meteoro* (LEMO FILHO *et al.*, 2013) e *MetFlow* (CARDOZO *et al.*, 2016). Além disso é necessário seguir a recomendação PROV da W3C (MOREAU *et al.*, 2014) para o registro da proveniência dos processos de transformação e anotação de dados.

Esta abordagem proposta se instancia como um *workflow* parametrizável que implementa análises preliminares e quantitativas discutidas na Seção 2 deste documento. O *CrossFlow* (Figura 3) é composto por quatro partes principais. Inicialmente ele faz a carga de dados das estações presentes do sistema *HidroWeb* da ANA (HIDROWEB, 2017), aplicam-se a ordenação dos dados, faz limpeza automática das séries, identifica e substitui-se os valores que extrapolam os limites físicos possíveis e efetua o preenchimento dos dados faltantes. Calculam-

se ainda as distâncias euclidianas entre as estações (por exemplo, raio = 30Km) baseado em parâmetros fornecidos pelo meteorologista.

Na segunda parte ocorre a validação cruzada do tipo *leave-one-out*. Nesta etapa, realizam-se os testes das amostras com os quatro métodos de interpolação (RL, PR, PRRL e IQD) para cada estação antes da seleção do método mais efetivo para o preenchimento de falhas. Com o intuito de avaliar o desempenho dos métodos são calculados os coeficientes de correlação (r) com critério mínimo de 0,7 para este coeficiente, e a raiz do erro médio quadrático (REMQ) (Tabela 1).

Na terceira parte, ocorre a assimilação de dados curados aos quais aplica-se o melhor dos quatro métodos de interpolação, preenchendo as falhas encontradas na *Etapa 1*. Nesta parte, também há o registro dos dados curados e dos dados de proveniência. Por fim, na quarta parte, ocorre a execução dos modelos meteorológicos adequados às pesquisas que utilizam os dados curados armazenados no banco de dados.

4. Avaliações Experimentais

Para avaliar a abordagem proposta, realizou-se uma série de experimentos com o *CrossFlow*, no qual utilizaram-se mais de 75 anos de séries de dados de dezenas de estações. Inicialmente, na *Etapa 1*, o *workflow* foi configurado para se conectar ao sistema *HidroWeb* e carregar todas as séries de dados no intervalo entre 1936 até 2013 das estações da região caracterizada na subseção 2.1 para o nosso banco de dados brutos. A seguir, executou-se a limpeza de cada uma das séries de dados de cada uma das estações (*Etapa 1*). Na Figura 4, são ilustrados todos os intervalos detectados sem falhas das 77 estações avaliadas (com raio de 30Km de separação entre elas).

A seguir, executou-se a limpeza de cada uma das séries de cada estação (*Etapa1*). Durante essa execução também foram elaborados mapas (apresentados na Figura 2) das estações alvo do estudo meteorológico que possuem séries com intervalos contínuos com o mínimo de 10 anos de médias mensais de precipitação e seus percentuais de falhas.



Figura 4 - Quantitativo de estações sem falhas no período 1936 até 2013.

Na *Etapa 2*, são executadas as atividades da validação cruzada para avaliar e identificar quais são os melhores métodos de preenchimento de falhas para a região e para o conjunto de estações estudadas. Na Tabela 1, são exibidos os índices da REMQ (em *mm*) para as 34 estações principais; sendo que em negrito são destacados os menores valores dos métodos de interpolação.

Ao analisar a Tabela 2 verificou-se que os métodos adotados para as análises quantitativas apresentadas na subseção 2.2 se mostram eficazes. Nesse estudo o método PR

obteve os menores índices da REMQ, avaliado satisfatoriamente em 73,5 % das estimativas. Os resultados obtidos mostram que o índice de correlação, em todas as estações avaliadas, foi superior a 0,7 para os quatro métodos utilizados. Pode-se destacar, em particular, a PR, que além dos menores índices de REQM, obteve correspondência superior a 0,9 em 55,9 % das regressões entre os valores estimados e os observados para a altura pluviométrica.

Tabela 1 - Resultados da raiz do erro médio quadrático (REMQ) em *mm* dos dados nas 34 estações principais.

| Código ANA | PR | IQD | PRRL | RL | Código ANA | PR | IQD | PRRL | RL |
|------------|--------------|--------------|--------------|--------|------------|--------------|--------------|--------------|-------|
| 2042027 | 35,19 | 35,68 | 34,49 | 47,00 | 2243005 | 35,76 | 43,51 | 38,98 | 38,79 |
| 2141003 | 36,24 | 40,84 | 42,29 | 41,02 | 2243006 | 35,16 | 39,42 | 36,02 | 48,27 |
| 2141006 | 30,53 | 33,39 | 32,19 | 37,89 | 2243007 | 46,13 | 49,41 | 46,99 | 51,67 |
| 2141007 | 28,79 | 29,70 | 33,41 | 41,36 | 2243008 | 32,11 | 37,17 | 34,06 | 47,28 |
| 2142014 | 34,79 | 39,53 | 34,37 | 37,60 | 2243009 | 59,80 | 55,47 | 59,16 | 57,71 |
| 2142016 | 46,65 | 44,92 | 45,52 | 46,07 | 2243010 | 36,88 | 40,43 | 37,21 | 46,60 |
| 2142022 | 37,75 | 40,94 | 40,79 | 48,97 | 2243011 | 31,21 | 35,35 | 32,79 | 46,16 |
| 2142058 | 33,76 | 34,15 | 35,03 | 42,69 | 2243012 | 27,63 | 26,43 | 27,22 | 31,15 |
| 2241004 | 46,72 | 54,09 | 54,91 | 142,10 | 2243013 | 34,11 | 38,68 | 38,54 | 44,63 |
| 2242001 | 39,59 | 41,17 | 39,56 | 47,44 | 2243014 | 28,74 | 28,29 | 28,65 | 30,99 |
| 2242002 | 47,00 | 51,61 | 46,13 | 47,17 | 2243015 | 34,51 | 38,24 | 37,61 | 53,49 |
| 2242015 | 45,35 | 60,09 | 61,85 | 54,11 | 2243016 | 35,45 | 36,00 | 36,42 | 52,77 |
| 2242017 | 45,64 | 57,95 | 57,74 | 76,93 | 2244038 | 59,29 | 63,76 | 66,54 | 61,82 |
| 2242018 | 47,94 | 50,72 | 50,10 | 62,33 | 2244039 | 83,04 | 82,51 | 81,94 | 97,26 |
| 2242019 | 35,58 | 37,73 | 37,14 | 64,05 | 2244041 | 39,26 | 63,26 | 57,75 | 52,45 |
| 2242023 | 53,70 | 74,08 | 70,29 | 109,34 | 2244043 | 43,01 | 55,66 | 48,42 | 54,56 |
| 2242029 | 43,22 | 42,73 | 42,91 | 49,89 | 2244046 | 52,57 | 60,62 | 70,77 | 61,14 |

Tabela 2 – Resumo dos resultados das estimativas

| Método | < REMQ | Estações com $r > 0,9$ |
|--------|--------|------------------------|
| PR | 25 | 19 |
| | 73,5% | 55,9% |
| IQD | 5 | 19 |
| | 14,7% | 55,9% |
| PRRL | 4 | 19 |
| | 11,8% | 55,9% |
| RL | 0 | 12 |
| | 0,0% | 35,3% |

Após a indicação do melhor método de preenchimento de falhas das séries, o *CrossFlow* é (re)parametrizado e de forma automática é novamente executado (*Etapa 3*). No entanto, desta vez as atividades de limpeza e validação não são executadas, pois o *CrossFlow* se conecta ao banco de dados curados e de proveniência, seleciona as 77 estações e aplica o melhor método (nesse estudo, o PR) em todas as séries de dados das estações previamente analisadas. Durante essa execução registram-se não só os dados brutos e curados, assim como os dados de proveniência retrospectiva (relativos às execuções do *workflow*, arquivos e seus parâmetros), isto é, para cada arquivo carregado, são correlacionados no repositório os dados originais (brutos), os dados transformados (curados), os dados de proveniência (método de interpolação aplicado), os agentes (usuários) envolvidos na execução do experimento, além das informações de *timestamp* sobre cada uma das transformações aplicadas. Por fim, na *Etapa 4*, o meteorologista pode utilizar os dados curados nos modelos meteorológicos que achar mais conveniente aos seus estudos (os modelos não estão apresentados neste texto).

Em vista do exposto, os primeiros resultados obtidos nos experimentos computacionais (já avaliados por equipes de meteorologistas) conduzem ao entendimento de que o uso da abordagem baseada em *workflows* e dados pluviométricos apoiados por validação cruzada tem

potencial para completar as falhas em longas séries de dados. Além disso, é ressaltada a capacidade dessa abordagem em realizar o processamento de grandes volumes de dados de forma automática, reduzir o tempo de análise por parte dos pesquisadores e gerar séries de dados de qualidade e anotadas por proveniência.

5. Trabalhos Relacionados

Existem abordagens na literatura capazes de tratar grandes volumes de dados meteorológicos e registrar a proveniência das execuções de *workflows* para verificar a qualidade destes dados. No entanto, nenhuma delas utiliza quatro métodos concomitantemente associados com a validação cruzada. Assim, apresentamos as abordagens mais semelhantes nessa seção.

Oliveira *et al.* (2010) propuseram um estudo comparativo entre os quatro métodos de interpolação para preenchimento de falhas cujo desenvolvimento e análise utilizou-se de planilhas eletrônicas. Os autores usaram dados de apenas seis estações por um período de apenas 22 anos e constataram que o método PR foi um dos melhores para o estudo de caso deles. Silva *et al.* (2014) avaliaram uma técnica de interpolação de dados baseada em imagens de satélite ao infravermelho em um sistema Web de controle de qualidade em dados de precipitação. Lemos Filho *et al.* (2013) desenvolveram um sistema Web de pré-processamento de dados pluviométricos e uso de proveniência associado ao método de RL para o preenchimento de falhas. Contudo, ressalta-se que nenhum dos autores citados acima utilizaram o paradigma dos *workflows* científicos.

De acordo com Guru *et al.* (2009) e Plale (2011) a utilização de *workflow* no ramo da Hidrologia representa diversas oportunidades de pesquisa. Por exemplo, os autores relatam as dificuldades relativas a integração e gerenciamento de dados e a necessidade da manutenção de dados brutos e curados para os estudos comparativos na área da qualidade. Saint e Murphy (2010) utilizam o sistema Kepler para automatizar o modelo hidrológico SWAT, que avalia o ciclo das águas em bacias hidrológicas da Austrália, no entanto, os autores não tecem maiores comentários sobre a etapa de carga e limpeza dos dados. Goodall *et al.* (2015) desenvolveram o sistema IRODS baseado em *workflows* para tratar dados meteorológicos de redes estadunidenses de operadoras distintas. Segundo os autores, esse esforço é necessário para assegurar não só a qualidade das pesquisas quanto para assegurar a reprodutibilidade dos experimentos.

6. Considerações Finais

Realizar o controle de qualidade de dados meteorológico não é uma tarefa trivial. Os dados brutos e interpolados são utilizados para previsões de tempo e clima e de trabalhos científicos. Portanto, assegurar sua qualidade, transparência e reprodutibilidade dos processos de produção de dados é de suma importância para o desenvolvimento de ações associadas ao planejamento urbano, industrial e agropecuário.

Diferentemente dos trabalhos disponíveis na literatura para a área, foi desenvolvido uma abordagem baseada em *workflows* científicos que utiliza validação cruzada baseada em quatro métodos de preenchimento que permitem aos meteorologistas transformar sistematicamente grandes volumes de dados inconsistentes em consistentes de qualidade e se beneficiarem da proveniência incorporada ao processo.

Por limitação de escopo este texto não se discutem as questões ligadas ao desempenho da abordagem nem as formas de recuperação de dados e de proveniência. Como trabalhos

futuros pretende-se adicionar um novo método de preenchimento de falhas (análise harmônica – série de Fourier). Além disso, planeja-se adicionar novos módulos que farão a análise de consistência de dados baseada no método de dupla massa e resíduo acumulado.

Agradecimentos

Os autores agradecem ao programa ao Programa de Educação Tutorial, ao FNDE, a Red BigDSSAgro CYTED e *Microsoft Azure Research* (CRM:0518152) pelos financiamentos.

Referências

- ANA. Orientações para Consistência de Dados Pluviométricos. Disponível em <<http://arquivos.ana.gov.br/infohidrologicas/cadastro/OrientacoesParaConsistenciaDadosPluviometricos-VersaoJul12.pdf>>. Acesso em 10 de fevereiro de 2017.
- Bertoni, J. C.; Tucci, C. E. M. Precipitação. In.: Hidrologia: Ciência e Aplicação, Org. Carlos E. M. Tucci, 2ª ed., Porto Alegre: Ed. Universidade/UFRGS. 2001.
- Callahan, S. P.; Freira, J.; Santos, E. *et al.* VisTrails: Visualization meets Data Management. In: Proc. SIGMOD 2006, p. 745-747, USA, 2006.
- Cardozo, F.; Tomaz, U. R.; Cruz, S. M. S. Avaliando Uma Estratégia Computacional Baseada Em Workflows Científicos Apoiados Por Placas Gráficas Genéricas. In: ERSI 2016 p. 66-73, Brasil.
- Goodall, J. L.; Essawy, B. T.; Xu, H.; Moore, R. W. Data processing workflows to support reproducible data-drive research in hydrology. AGU Fall Meeting, 2015.
- Guru, S. M.; Kearney, M.; Fitch, P.; Peters, C. Challenges in using scientific workflow tools in the hydrology domain. 18th IMACS, Int. Congress on Modelling and Simulation, Austrália, p. 3514-3520, 2009.
- HidroWeb - Sistema HidroWeb. Disponível em: <<http://hidroweb.ana.gov.br>>. Acesso em 14 de jan de 2017.
- Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th Int. Joint Conf. on Artificial Intelligence – v. 2. San Francisco, CA, USA, p. 1137-1143, 1995.
- Lemos Filho, G. R.; Oliveira, E. S.; Lyra, G. B.; Cruz, S.M.S. Assimilação, Controle de Qualidade e Análise de Dados de Meteorológicos Apoiados por Proveniência, VII e-Science Workshop, XXXIII CSBC. 2013.
- Marengo, J. A. Mudanças climáticas, condições meteorológicas extremas e eventos climáticos no Brasil, p. 4-19. In: Marengo, J. A., Schaeffer, R.; Zee, D. M. W.; Pinto, H. S. Mudanças climáticas e eventos extremos no Brasil. p. 1-76. Fundação Brasileira para o Desenvolvimento Sustentável, FBDS. 2009.
- Murphy S.; Saint K.; End-to-End Workflows for Coupled Climate and Hydrological Modeling. In: Int. Env. Modelling and Software Society. 8 pp. 2010.
- Oliveira, L. F. C.; Fioreze, A. P.; Medeiros, A. M. M.; Silva, M. A. S. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. In: Rev. Bras. Eng. Agrícola e Ambiental, 14(11) 1186-1192. 2010.
- Moreau, L.; Missier, P.; Belhajjame, K. *et al.* PROV-DM: The PROV Data Model. W3C Recommendation REC-prov-dm-20130430, 2013.
- Plale, B. The Challenges and Opportunities of Workflow Systems in Environmental Research. 2011. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.298.8237>
- Silva, F. D. S.; Ramos, R. M.; Costa, R. L.; Azevedo, P. V. Sistema de controle de qualidade para dados diários de variáveis meteorológicas. Rev. Bras. de Geografia Física, vol, 07, n. 05, 2014.